Identification and *in silico* analysis of the *Citrus* HSP70 molecular chaperone gene family

Luciano G. Fietto¹*, Maximiller D.L. Costa¹*, Cosme D. Cruz², Alessandra A. Souza³, Marcos A. Machado³ and Elizabeth P.B. Fontes¹

¹Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, MG, Brazil.
²Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, Brazil.
³Centro APTA Citros Sylvio Moreira, Instituto Agronômico de Campinas, Cordeirópolis, SP, Brazil.

Abstract

The completion of the genome sequencing of the *Arabidopsis thaliana* model system provided a powerful molecular tool for comparative analysis of gene families present in the genome of economically relevant plant species. In this investigation, we used the sequences of the *Arabidopsis* Hsp70 gene family to identify and annotate the *Citrus* Hsp70 genes represented in the CitEST database. Based on sequence comparison analysis, we identified 18 clusters that were further divided into 5 subgroups encoding four mitochondrial mtHsp70s, three plastid csHsp70s, one ER luminal Hsp70 BiP, two HSP110/SSE-related proteins and eight cytosolic Hsp/Hsc70s. We also analyzed the expression profile by digital Northern of each Hsp70 transcript in different organs and in response to stress conditions. The EST database revealed a distinct population distribution of Hsp70 ESTs among isoforms and across the organs surveyed. The Hsp70-5 isoform was highly expressed in seeds, whereas BiP, mitochondrial and plastid Hsp70 mRNAs displayed a similar expression profile in the organs analyzed, and were predominantly represented in flowers. Distinct Hsp70 mRNAs were also differentially expressed during *Xylella* infection and *Citrus tristeza* viral infection as well as during water deficit. This *in silico* study sets the groundwork for future investigations to fully characterize functionally the *Citrus* Hsp70 family and underscores the relevance of Hsp70s in response to abiotic and biotic stresses in *Citrus*.

Key words: chaperones, digital northern, heat shock proteins, genome, expressed sequence tags.

Received: July 21, 2006; Accepted: November 29, 2007.

Introduction

The Hsp70 family of stress-related proteins comprises a series of highly conserved ATPase molecular chaperones that are distributed ubiquitously in all prokaryotes and in the major subcellular compartments of eukaryotes. They have been implicated as key mediators of the correct folding of newly synthesized proteins, translocation of proteins across the membranes, assembly and disassembly of multimeric protein complexes, as well as disassembly of stress-induced protein aggregates and targeting of abnormal proteins to degradation (Hartl, 1996; Miernyk, 1997; Frydman, 2001). Hsp70s were first described as heat shock proteins induced by a rapid increase in temperature, but subsequently were recognized as general stress-related proteins that respond to a variety of stress conditions and are also required to assist proper folding of proteins under normal temperature. In general, the constitutively expressed members of the Hsp70 family, referred to as Hsc70 (heat shock cognate), assist in the proper folding of newly synthesized proteins and the translocation of proteins into organelles, whereas the stress-induced members are involved in the re-folding and degradation of misfolded proteins under adverse environmental conditions (Hartl, 1996; Miernyk, 1997; Frydman, 2001). Furthermore, eukaryotic cytosolic Hsp70s also play a role in controlling the activity of regulatory proteins, such as the transcriptional repression of HSF (Heat shock factor) (Shi et al., 1998) and the modulation of steroid receptor function (Morishima et al., 2000).

The Hsp70s are structurally organized into a highly conserved ATPase domain at their N-terminal portion and a peptide binding domain at their C-terminus. Successive cycles of peptide binding and release are coupled to the activity of ATPase that requires the participation of co-chaperones, such as DnaJ/Hsp40 and GPrE (Bukau and Horwich, 1998).
Among Hsps, the Hsp70 family encompasses the most conserved members exhibiting about 50% sequence identity between representatives of such diverse organisms as \textit{E. coli} (DnaK) and eukaryotes (Hsp70s). In the same eukaryotic organism, Hsp70s from different compartments maintain high conservation of sequence and, in general, the function of the distinct isoforms may be predicted according to their subcellular localization.

In plants, several Hsp70 genes have been identified in the last decade (Boston et al., 1996; Guy and Li, 1998; Vierling, 1991) but a global picture of the plant Hsp70 family has emerged with the completion of the genome sequencing of \textit{Arabidopsis thaliana} (Sung et al., 2001a; 2001b; Borges et al., 2001; Cagliari et al., 2005). The fully sequenced Arabidopsis genome contains 18 identified Hsp70 genes, which, based on subcellular localization and sequence identity, have been categorized into five subfamilies, the cytosolic Hsp/Hsc70s, the plastid cpHsc70s, the mitochondrial mtHsc70s, the endoplasmic reticulum (ER) luminal BiPs and the subfamily of Hsp91. The plant Hsp/Hsc70 subfamily is highly conserved encompassing homologous genes that share more than 75% sequence identity. The members of both plastid and mitochondrial Hsc70 subfamilies are most related to DnaK, the bacterial Hsp70 homolog, whereas the ER luminal BiPs and Hsp70s from other compartments are closely related to the cytosolic Hsp70s (Wimmer et al., 1997). Of the 18 genes in Arabidopsis, 14 Hsp70 genes are \textit{E. coli} DnaK homologs and four (4) genes are HSP110/SSE homologs (Lin et al., 2001).

Although the Hsp70 family maintains highly conserved structural and functional features across evolutionarily distant organisms, some structural characteristics are unique to plants. These include the presence of multiple BiP genes in the genome of some plant species and the presence of a predictive localization motif for organelar Hsp70 proteins at the C-terminus of the plant genes (Sung et al., 2001a). In fact, plant BiP subfamilies have been identified in tobacco comprising 5 genes, in soybean (4 genes), in Arabidopsis (3 genes), in Eucalyptus (5 genes) and in maize with at least 2 genes (Denecke et al., 1991; Wrobel et al., 1997; Cascardo et al., 2000; 2001; Fontes et al., 1991; Sung et al., 2001b; Cagliari et al., 2005). As for plant Hsp70 localization motifs, they are contained in the unique and highly conserved C-terminal region of each Hsp70 from the same subgroup. They correspond to the EEVD motif for the cytosolic Hsp70, the HDEL sequence for the ER-resident Hsp70s (BiPs), the PEAEEYAKK sequence for mitochondrial Hsp70s and the PEGVIDADFTDSK motif for the chloroplast-localized homologs (Guy and Li, 1998). While the HDEL sequence functions as an ER retention signal, the other conserved sequences appear to act as an interaction site for co-chaperones that are specific to the organelles in which the Hsp70s are located.

Biotic and abiotic stresses, such as salinity, temperature extremes, oxidative stress, fungal and virus infection, represent major constraints for the agriculture and are responsible for 50% of crop loss worldwide (Wang et al., 2003). The elucidation of the mechanisms of plant response to stresses and how it acquires tolerance is absolutely relevant for breeding programs. Since Hsp70 plays a crucial role in protecting against stresses and in the re-establishment of cellular homeostasis, the characterization of the plant Hsp70 family and its complex regulation is likely to lead to new strategies to enhance crop tolerance to environmental stress. Consistent with this observation, overexpression of Hsp70 has been shown to be positively correlated to the acquisition of thermotolerance and to an increase in tolerance to water deficit and salt stress (Lee and Schöffl, 1996; Leborgne-Castel et al., 1999; Alvim et al., 2001; Sung and Guy, 2003). However, the underlying mechanisms of the Hsp70-mediated multiple cross tolerance and the role of Hsp70 during stress conditions have not been completely deciphered yet. Given the differential regulation of the plant Hsp70 gene family and the likely functional diversity of its members, a comprehensive characterization of the Hsp70 family in a plant species is required to understand how Hsp70s contribute to cell function and protection against specific environmental stressors. Here, we used an \textit{in silico} approach to identify the \textit{Citrus} Hsp70 genes and characterize the expression pattern of this family. We searched the CitEST database for tissue-specific abundance of Hsp70 mRNAs in libraries from different tissues and under different stress conditions. We identified at least 18 Hsp70 sequences from \textit{Citrus} and showed that the Hsp70 isoforms display a diverse pattern of tissue-specific expression and are differentially regulated in response to biotic and abiotic stresses.

**Material and Methods**

In this investigation, we analyzed the CitEST database (http://biotecnologia.centrodecitricultura.br) to identify clusters that potentially encode Hsp70s. For the sequence analysis we used the BLAST program (Altschul et al., 1997) and we searched for each \textit{Citrus} HSP70 gene using an \textit{Arabidopsis} representative sequence of each Hsp70 subgroup (AT5G42020.1, AT1G6030.1, AT5G49910.1, AT1G79930.1, AT3G12580.1, AT4G37910.1) as prototypes. The searches were performed using tBLASTn with a score cutoff of E\textsuperscript{-60} (expected value). Partially overlapping ESTs were assembled into contigs and each derived cluster was considered a potential gene. The tentative consensus DNA sequences derived from the overlapping ESTs were used as prototypes for a new search cycle, which was performed using BLASTn with a cutoff of E\textsuperscript{-80} (expected value).

The possible genes were translated and the deduced protein sequences were analyzed. The percentage of amino acid sequence identity was determined using the ClustalW
program (Thompson et al., 1994) or LALIGN program (Pearson, 1990). To generate the Dendogram, the sequences alignment in phylip format were obtained with ClustalW and the tree was generated using the Drawgram Program (Thompson et al., 1994). The subcellular localization of HSP70 genes was predicted by using the PREDOTAR software (http://urgi.infobiogen.fr/predotar/predotar.html). To evaluate gene expression in different libraries, the in silico transcriptional profiling or digital Northern analysis was performed by counting and dividing the number of sequenced ESTs for each annotated Hsp70 gene by the whole sequenced EST population of each library. The relative frequency of each gene was considered over a library size corrected to 10,000 ESTs and the differential gene expression was statistically analyzed as described by Audic and Claverie (1997).

Results and Discussion

The Citrus Hsp70 multigene family comprises at least five subfamilies

The analysis of the CitEST database, which was generated through sequencing of ESTs from cDNA libraries prepared from different tissues and species of Citrus (Citrus sinensis, Citrus reticulata, Citrus limonia, Citrus aurantium and Citrus aurantifolia), suggests the existence of at least 18 Hsp70 genes in the Citrus genome (Table 1). As the libraries displayed great variation in size and tissue representation among the species, the generated data were from a collective and global analysis. While this strategy produced statistically meaningful data, it did not rule out the possibility that the highly conserved isoforms were allelic variations of the same gene present in distinct Citrus species. Based on sequence comparison (Figure 1), prediction of subcellular localization (Table 1) and the presence of characteristic domains of the different HSP70 subgroups from other plant species (Sung et al., 2001a), the Citrus Hsp70-encoded proteins were grouped into five distinct classes or subfamilies. The four major subfamilies are represented by eight cytosolic Hsp/Hsc70, one ER-resident BiP, four mitochondrial mtHsp70 and three csHsp70 from the chloroplast. The fifth subgroup is represented by two Hsp110/SSE-related high molecular weight Hsp70s (Table 1). Of the 18 probable genes in Citrus, only four represent full-length sequence contigs, Hsc70-5, mtHsp70-1, csHsp70-1 and Hsp70-1 (Table 1), and in the BiP-1 sequence, a few amino acid residues are missing at the N-terminal region, as compared to the Lycopersicon esculentum BiP deduced protein sequence (Genbank protein accession number, gi:1346172).

The Citrus Hsp70 categorization was consistent with the dendogram shown in Fig. 1, which was constructed by comparing the Citrus full-length Hsp70 sequences with the Arabidopsis counterpart of each subgroup (http://www.arabidopsis.org). Each representative of the Citrus Hsp70

<table>
<thead>
<tr>
<th>Hsp70 member</th>
<th>Hsp70 gene family</th>
<th>Best Hit</th>
<th>E-value</th>
<th>N. of reads</th>
<th>aa</th>
<th>Full length EIVDP seed</th>
<th>HDEL1 seed</th>
<th>Predator</th>
<th>PEAEYEEK seed</th>
<th>PEGDYKADTTDY seed</th>
<th>PEGDYKADTTDY seed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsc70-1</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>0.0</td>
<td>2-13</td>
<td>4</td>
<td>356/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-2</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>2e-16</td>
<td>2-16</td>
<td>6</td>
<td>294/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-3</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>1e-14</td>
<td>2-14</td>
<td>6</td>
<td>224/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-4</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>1e-20</td>
<td>2-20</td>
<td>12</td>
<td>685/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-5</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>0.0</td>
<td>3-31</td>
<td>21</td>
<td>209/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-6</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>5e-129</td>
<td>2-129</td>
<td>13</td>
<td>295/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-7</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>9e-124</td>
<td>2-124</td>
<td>10</td>
<td>280/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-8</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>3e-89</td>
<td>2-89</td>
<td>3</td>
<td>204/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-9</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>3e-86</td>
<td>2-86</td>
<td>8</td>
<td>235/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-10</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>0.0</td>
<td>3-37</td>
<td>10</td>
<td>244/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-11</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>1e-15</td>
<td>2-15</td>
<td>10</td>
<td>244/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsc70-12</td>
<td>Citrus Hsp70</td>
<td>Hsc70</td>
<td>1e-115</td>
<td>2-115</td>
<td>10</td>
<td>244/647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1C-terminal amino acid signature for different subgroups of the Hsp70 protein. 2Number of amino acid residues covered by the cluster. 3Full-length protein. 4ER: endoplasmic reticulum; ND: not determined.
Plant Hsp70s form a family of stress-related proteins that has been further subdivided into categories based on amino acid sequence homology and the subcellular locations of the proteins (Sung et al., 2001a). Although the size and complexity of the Citrus Hsp70 family described here closely resemble those described for this family in other plant species (Guy and Li, 1998; Sung et al., 2001b; Borges et al., 2001; Cagliari et al., 2005), the ultimate understanding of the entire complement of Citrus Hsp70 genes depends on the identification of the full-length sequences of all representatives of the family.

Tissue expression pattern of the Citrus Hsp70 genes

For insights into the diverse roles of Hsp70s and to determine whether the different Citrus isoforms are functionally redundant or divergent, we next conducted a digital northern analysis on the expression pattern of these identified Citrus genes in different organs and in response to abiotic and biotic stresses. We analyzed the frequency of the ESTs of the different Hsp70 isoforms and only those Hsp70s that were represented by at least five reads were considered (Table 1). Figure 2 shows a comparison of relative mRNA prevalence across representative organs. Except for the Hsc70-5 transcript that displayed a statistically significant prevalence in seeds as compared to the other organs analyzed, the transcripts of all the other isoforms were not detected in mature seeds. Similar transcriptional profiling has been observed for the members of the Arabidopsis family in which a cytosolic Hsp70 isoform (Hsc70-12) is predominantly expressed in seeds, whereas the luminal and organellar Hsp70 transcripts are barely detected in mature dry seeds (Sung et al., 2001b). The decline in the expression of Hsp70 genes at later stages of seed development may be offset by the increased expression of other chaperones as in the Citrus cytosolic Hsc70-5 isoform. Alternatively or additionally, it may implicate a minor role of Hsp70 genes in mature seeds. The expression of the luminal Bip has been monitored during seed development in several plant species (Carolina et al., 2003). In general, an increase in Bip expression during seed development parallels the accumulation of seed storage proteins, decreasing rapidly toward the end of seed maturation. This developmental regulation of Bip during seed maturation is coordinated with the cellular secretory activity of seeds and implies a
relevant role of BiP in newly-synthesized storage protein folding.

Although we did not observe statistically significant differences in the prevalence of other Hsp70 transcripts across the range of organs surveyed, BiP-1 and Hsc70-6 as well as csHsp70-1 and mtHsp70-1 displayed similar expression pattern in all organs, but their transcript levels predominated in flower. Other Hsp70 isoforms, such as csHsp70-3, Hsc70-4 and Hsp70-2, were dramatically underrepresented in all organs. Thus, considering the level of transcript abundance, several Hsp70 genes display a broad expression pattern consistent with a generalized housekeeping function, whereas other Hsp70 mRNAs are rare and may represent temperature- or other stress-induced forms. The tissue-specific libraries have representatives of different species (Citrus sinensis, Citrus reticulata, Citrus limonia, Citrus aurantium and Citrus aurantifolia) and although they were analyzed collectively, as required to obtain statistically meaningful data, we observed a differential expression of the Citrus Hsp70 family among the species (data not shown). The physiological implications of this finding are unknown and whether the different species display distinct tolerance levels to environmental stresses as a function of Hsp70 expression remains to be determined.

**Hsp70 expression during biotic and abiotic stress**

We were particularly interested in evaluating the expression of the Citrus Hsp70 gene family in response to physiological stresses that affect drastically the Citrus productivity, such as Xylella infection, viral infection and water deficit. The expression pattern of Hsp70 genes from Citrus sinensis and Citrus reticulata during the process of xylella infection is shown in Figure 3A. BiP-1, csHps70-3 and mtHsp70-1 were strongly induced during Xylella infection and their response was paralleled by the opposite trend with cytosolic Hsp70 representatives, in which the transcript levels declined upon infection. While induction of the csHps70-3 and mtHsp70-1 may be a direct result of oxidative stress caused by the bacterial infection, BiP induction may reflect the activation of the unfolded protein response (UPR) pathway. This signaling pathway, which is activated by ER stressors that cause the accumulation of unfolded proteins in the endoplasmic reticulum (ER) lu-

Figure 2 - Expression of the Citrus Hsp70 genes in the different tissues of Citrus. The expression pattern was calculated as indicated in the materials and methods and normalized according the library. An asterisk (*) indicates an 80% probability of increased Hsc70-5 expression in seeds as compared to the other organs analyzed. A plus sign (+) indicates more than 95% probability of increased expression of Hsc70-5 in seeds as compared to the seed expression of the other Hsp70 members analyzed.

Figure 3 - Expression of the Citrus Hsp70 genes in response to different stresses. (A) expression in response to Xylella infection. An asterisk (*) indicates an 80% probability of differential expression among the different infection stages. (B) Expression pattern in response to Citrus tristeza virus (CTV) infection. The asterisk (*) indicates an 80% probability for differential expression between infected and uninfected plants (C) Expression pattern in the Citrus root under drought stress. A plus sign (+) indicates more than 95% probability of increased expression in stressed roots as compared to unstressed roots.
men, triggers the coordinated up-regulation of a set of folding enzymes and ER molecular chaperones, including BiP and the Ca\(^{2+}\)-binding protein calreticulin (Kauffman, 1999). In order to determine whether Xylella infection provoked ER stress and hence activation of the UPR, we searched for the calreticulin ortholog in the Citrestricula virus-infected Poincirus trifoliata leaves (Fig. 3B). The reason for this selective response to viral infection is not clear, but it may be related to an enhanced demand of chaperone function in the mitochondria of infected cells, linked to an exhaustive use of ATP-demanding biosynthetic pathways for replication and expression of the viral genome. Alternatively, it may represent a specific protection mechanism of plant cells against viral infection.

We also searched for Hsp70 transcript abundance in EST libraries prepared from water deficit-stressed root mRNA (Fig. 3C). Both mtHsp70-1 and Hsc70-5 mRNAs are selectively up-regulated in response to water dehydration. Induction of mtHsp70-1 may reflect a protective mechanism of plant cells against the deleterious effects caused by the excessive accumulation of reactive oxygen species (ROS) under dehydration (Bartels, 2001). The induction of Hsc70-5 under water deficit may be particularly and biologically relevant as it is also induced during late seed maturation (Fig. 2) when seed dehydration takes place under a desiccation-tolerant regime that retains seed viability. In fact, the pattern of Hsc70-5 expression resembles the pattern of Hsc70-5 expression elicited by Xylella infection, only the mtHsp70-1 transcript displayed a statistically significant increased level in the Citrus tristeteza virus-infected Poncirus trifoliata leaves (data not shown) that is considered resistant to the bacterium. In contrast to the Hsp70 expression pattern elicited by Xylella infection, only the mtHsp70-1 transcript displayed a statistically significant increased level in the Citrus tristeteza virus-infected Poncirus trifoliata leaves (Fig. 3B). The reason for this selective response to viral infection is not clear, but it may be related to an enhanced demand of chaperone function in the mitochondria of infected cells, linked to an exhaustive use of ATP-demanding biosynthetic pathways for replication and expression of the viral genome. Alternatively, it may represent a specific protection mechanism of plant cells against viral infection.

We also searched for Hsp70 transcript abundance in EST libraries prepared from water deficit-stressed root mRNA (Fig. 3C). Both mtHsp70-1 and Hsc70-5 mRNAs are selectively up-regulated in response to water dehydration. Induction of mtHsp70-1 may reflect a protective mechanism of plant cells against the deleterious effects caused by the excessive accumulation of reactive oxygen species (ROS) under dehydration (Bartels, 2001). The induction of Hsc70-5 under water deficit may be particularly and biologically relevant as it is also induced during late seed maturation (Fig. 2) when seed dehydration takes place under a desiccation-tolerant regime that retains seed viability. In fact, the pattern of Hsc70-5 expression resembles that of Lea (late embryogenesis abundant) and small Hsps are believed to be involved (Ingram and Bartels, 1996; Wehmeyer and Vierling, 2000). The accumulation of Lea proteins and smHsp occurs during seed maturation, desiccation and increases in vegetative tissue when plants are exposed to water deficit. The molecular mechanisms that confer desiccation tolerance and allow seeds to be dried without loss of viability are incompletely defined, although Lea proteins and small Hsps are believed to be involved (Ingram and Bartels, 1996; Wehmeyer and Vierling, 2000). Whether Citrus Hsc70-5 plays a specific role in seed desiccation tolerance or in the general protection of cellular components during plant cell dehydration will require further investigations.

Acknowledgments

This research was supported by CNPq grants 50.6119/2004-1 and 470787/2004-0 (to E.P.B.F.) and FAPEMIG grants EDT 560/05 and CBB 1493/05 (to E.P.B.F.). M.D.L.C. is supported by a CAPES graduate fellowship.

References


Internet Resources

Citrus Biotecnology Laboratory (CitEST), http://biotecnologia.centrodecitricultura.br (April 15, 2006)

A prediction service for identifying putative N-terminal targeting sequences (PREDOTAR), http://urgi.infobiogen.fr/predotar/predotar.html (April 17, 2006)


Associate Editor: Raquel Luciana Boscariol-Camargo